NONLINEAR MAPPING DESCRIBING LEARNING PROCESS HAVING MEMORY

Sang Seok Lee and Hiroaki Hara
Department of Information Science, Graduate School of Tohoku University, Sendai 980, Japan
Tamura Yosiyasu
The Institute of Statistical Mathematics, Minato-Ku, Tokyo 106, Japan

A nonlinear mapping model for learning and memory is proposed. The learning is a process that visual patterns presented in an information source(IS) cause selective activations of units composing a random network(RN). The process is a nonlinear map from the visual patterns onto the sets of the selective activations of the units. The aggregation is a set of the clusters in which the units are interconnected strongly with a coupling line, while the other units, not belonging to the clusters, are interconnected loosely. During a lapse of time, the states are assumed to be modified by two effects for the firing units; a rehearsal effect and a forgetting effect. The affected states show serial position curves(SPC) on which the points are arranged by a sequential list of the visual patterns presented.

We present the visual patterns, A_k $(k = 1, 2, \cdots, n)$, specified with occurrence probabilities $P(A_k)$ in the randomly arranged IS. Let X be an original element defined by

$$X = \begin{pmatrix} A_1, & A_2, & \cdots, & A_n \\ P(A_1), & P(A_2), & \cdots, & P(A_n) \end{pmatrix}, \tag{1}$$

where $\sum_{k=1}^{n} P(A_k) = 1$. The mapping is decomposed by successive ones $\varphi^{(k)}$ $(k = 1, 2, \cdots, n)$. The aggregation of the clusters yield a potential denoted by $y^{(k)}$. To define $y^{(k)}$, let $M(A_k)$'s be clusters and c_k be the number of the clusters $M(A_k)$. Now we introduce a normalized number of $M(A_k)$, that has a probability ρ_k, i.e. $\rho_k = c_k / \sum_{k=1}^{n} c_k$. Then the mapping is formally defined as follow

$$\varphi^{(k)} \ : \ X^{(k)} = \begin{pmatrix} A_k \\ P(A_k) \end{pmatrix} \longrightarrow Y^{(k)} = \begin{pmatrix} M(A_k) \\ \rho_k \end{pmatrix}. \tag{2}$$

In the mapping $\varphi^{(k)}$'s, we have not to take account of a time delay of $y^{(k)}$ explicitly. To evaluate the time delay, we divide the potential $y^{(k)}$ into its components $y_j^{(k)}$ and we regard $y^{(k)}$ as a vector potential. Expressed differently, we associate the elements $Y^{(k)}$'s $(k = 1, 2, \cdots, n)$ representing the aggregations with the vector potential $y^{(k)}$'s ;

$$\begin{aligned} y^{(k)} \ &= \ (y_1^{(k)}, y_2^{(k)}, \cdots, y_k^{(k)}, 0, \cdots, 0), \\ \longleftrightarrow Y^{(k)} \ &\equiv \ E^{(k-1)} \cdot Y^{(k-1)} + \varphi^{(k)}(X^{(k)}), \end{aligned} \tag{3}$$

where $E^{(k-1)}$ is a $(k-1) \times (k-1)$ matrix specifying a correlation between the mappings $\varphi^{(k-1)}$'s and the j-th components of $y^{(k)}$ are given by

$$y_j^{(k)} = \begin{cases} \displaystyle\sum_{j=1}^{k-1} e_{ij}^{(k-1)} y_j^{(k-1)} & (1 \le i \le k-1, 1 \le j \le k-1) \\ y_k^{(k)} & (j = k) \end{cases}, \tag{4}$$

where $e_{ij}^{(k)}$'s are the matrix elements. With the aid of the matrix element $e_{ij}^{(k)}$, we characterize the correlation by the effective coupling lines. During a lapse of time, the effective coupling lines between the clusters will be changed.

To evaluate the memory effects arising from the correlations, we suppose that $e_{ij}^{(k)}$'s are characterized by the two effects,

$$e_{ij}^{(k)} = R_{ij}^{(k)} + D_{ij}^{(k)}, \tag{5}$$

where $R_{ij}^{(k)}$ and $D_{ij}^{(k)}$ represent a rehearsal effect and a forgetting(decaying) effect, respectively.

We assume exponential dependences for $R_{ij}^{(k)}$ and $D_{ij}^{(k)}$;

$$R_{ij}^{(k)} = R_{ij}^{(0)} exp(-r_0(ij)\ \{k-(i+j)/2\})^2, \tag{6}$$

$$D_{ij}^{(k)} = D_{ij}^{(0)} exp(-d_0(ij)\ \{k-(i+j)/2\})^2, \tag{7}$$

for $i, j = 1, 2, \cdots, k-1$ and $k = 1, 2, \cdots n$, in the numerical calculation.

For our purpose it is sufficient to know that the notations $R_{ij}^{(0)}$, $D_{ij}^{(0)}$, $r_0(ij)\ (< 0)$, $d_0(ij)\ (> 0)$ denote some values of constant.

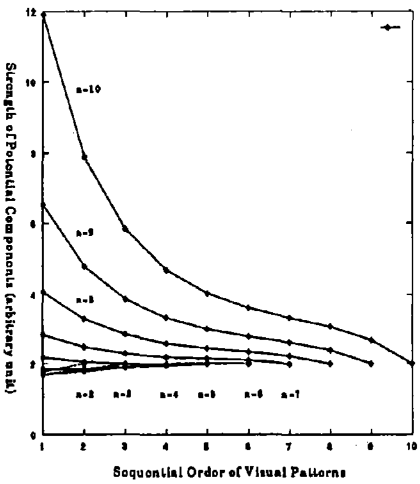

Figure 1: The behaviors of the errors show serial position curves(SPC). In the numerical calculations, we put $R_{ii}^{(0)} = 1.0$, $R_{ij}^{(0)} = 0.04$, $D_{ii}^{(0)} = -0.1$, $D_{ij}^{(0)} = -0.01$, and we also use the constants given as $r_0(ii) = 0.1$, $r_0(ij) = -0.005$, $d_0(ii) = 0.1$, $d_0(ij) = -0.005$, respectively.

Fig. 1 shows behaviors of variation for potential components, by which the errors would be interpreted.

As a result, we gave a nonlinear mapping model for memory and learning. The learning was a process expressed by the nonlinear map from IS onto RN. States of the RN were determined by the selective activations. Furthermore, we introduced aggregations of the clusters. During a lapse of time, the coupling line between the units or the effective coupling line between the clusters have changed, and eventually the changes have caused errors in a retrieval task. As curves showing the errors we obtained serial position curves(SPC).